

WHAT IS CLAIMED IS:

1. An information extracting apparatus for extracting designated information from a document group having a hypertext structure in which documents are mutually related by link information, comprising:

a start point address designating unit which designates an address of the document serving as a start point where said information is extracted; and

an extracting unit which extracts said information from the target document designated by said start point designating unit and, if said information could not be extracted from said target document, extracts said information from a related document of said target document on the basis of the address of said document.

2. The apparatus according to claim 1, further comprising:

an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

3. The apparatus according to claim 1, further comprising:

a maximum link depth designating unit which designates a maximum link depth; and

an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

4. The apparatus according to claim 3, further comprising:  
an extracting unit which discriminates an internal link and an  
external link on the basis of the document address of the related document and  
excludes the documents of the external link from the targets of the information  
extraction.

5. The apparatus according to claim 3, further comprising:  
an extracting unit which executes the information extracting  
process in order of the document in which a value of the link depth is small.

6. The apparatus according to claim 5, further comprising:  
an extracting unit which discriminates an internal link and an  
external link on the basis of the document address of the related document and  
excludes the documents of the external link from the targets of the information  
extraction.

7. The apparatus according to claim 1, wherein said related  
document includes at least one of a link destination document, a link source  
document, and an upper document of the target document.

8. The apparatus according to claim 7, wherein said upper document  
is at least either a document of a specific name existing in a one-upper  
directory of the target document or a link source document existing in the one-  
upper directory.

9. The apparatus according to claim 1, further comprising:

a category designating unit which designates a category of the information to be extracted; and

an extracting unit which extracts the information corresponding to said category from the target document designated by said start point address designating unit and, if the information corresponding to said category could not be extracted from said target document, extracts said information from the related document of said target document on the basis of the address of said document.

10. The apparatus according to claim 9, further comprising:

an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

11. The apparatus according to claim 9, further comprising:

a maximum link depth designating unit which designates a maximum link depth; and

an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

12. The apparatus according to claim 11, further comprising:

an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information

extraction.

13. The apparatus according to claim 11, further comprising:  
an extracting unit which executes the information extracting  
5 process in order of the document in which a value of the link depth is small.

14. The apparatus according to claim 13, further comprising:  
an extracting unit which discriminates an internal link and an  
external link on the basis of the document address of the related document and  
10 excludes the documents of the external link from the targets of the information  
extraction.

15. The apparatus according to claim 9, wherein said related  
document includes at least one of a link destination document, a link source  
15 document, and an upper document of the target document.

16. The apparatus according to claim 15, wherein said upper  
document is at least either a document of a specific name existing in a one-  
upper directory of the target document or a link source document existing in  
20 the one-upper directory.

17. The apparatus according to claim 9, further comprising:  
a category layer specifying unit in which the category of the  
information to be extracted is expressed by a layer structure;  
25 an extracting unit which, in the case where only an extraction  
result of a lower layer in said layer structure exists and an extraction result of  
an upper layer is missing as a result of the extraction of the information

corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result of said lower layer from the related document of said target document; and

5                   a processing unit which outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer.

18.           The apparatus according to claim 17, further comprising:

10                   a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule.

15           19.           The apparatus according to claim 17, further comprising:

                  a processing unit which forms the character string of the processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure.

20           20.           The apparatus according to claim 19, further comprising:

                  a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule.

25

21.           The apparatus according to claim 17, further comprising:

an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

5

22. The apparatus according to claim 17, further comprising:  
a maximum link depth designating unit which designates a maximum link depth; and

10

an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

15

23. The apparatus according to claim 22, further comprising:  
an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

20

24. The apparatus according to claim 22, further comprising:  
an extracting unit which executes the information extracting process in order of the document in which a value of the link depth is small.

25

25. The apparatus according to claim 24, further comprising:  
an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information

extraction.

26. The apparatus according to claim 17, wherein said related document includes at least one of a link destination document, a link source document, and an upper document of the target document.

27. The apparatus according to claim 26, wherein said upper document is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory.

28. The apparatus according to claim 17, further comprising:  
an extracting unit which, in the case where the extraction result is separated into a plurality of character strings of the extraction result of the lower layer and the extraction result of the upper layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, outputs said plurality of character strings as an extraction result of the lower layer and an extraction result of the upper layer.

29. The apparatus according to claim 28, further comprising:  
a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule.

30. The apparatus according to claim 28, further comprising:

a processing unit which forms the character string of the processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure.

5

31. The apparatus according to claim 30, further comprising:

a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule.

10

32. The apparatus according to claim 28, further comprising:

an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

15

33. The apparatus according to claim 28, further comprising:

a maximum link depth designating unit which designates a maximum link depth; and

20

an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

25

34. The apparatus according to claim 33, further comprising:

an extracting unit which discriminates an internal link and an



external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

5        35.        The apparatus according to claim 33, further comprising:  
                 an extracting unit which executes the information extracting  
process in order of the document in which a value of the link depth is small.

10       36.        The apparatus according to claim 35, further comprising:  
                 an extracting unit which discriminates an internal link and an  
external link on the basis of the document address of the related document and  
excludes the documents of the external link from the targets of the information  
extraction.

15       37.        The apparatus according to claim 28, wherein said related  
document includes at least one of a link destination document, a link source  
document, and an upper document of the target document.

20       38.        The apparatus according to claim 37, wherein said upper  
document is at least either a document of a specific name existing in a one-  
upper directory of the target document or a link source document existing in  
the one-upper directory.

25       39.        An information extracting apparatus for extracting designated  
information from a document group having a hypertext structure in which  
documents are mutually related by link information, comprising:  
                 an extracting unit which extracts target information from said

document group and, in the case where addition or updating of a document occurs for said document group, executes an extracting process to which such addition or updating is reflected each time said addition or updating occurs, and outputs an extraction result including said target information and its document address;

an extraction result storing unit which stores the extraction result from said extracting unit as extraction result information;

a start point address designating unit which designates an address of a document serving as a start point where said designated information is extracted; and

a searching unit which extracts information from the document of the document address designated by said start point address designating unit and its related document with reference to the extraction result information in said extraction result storing unit.

40. The apparatus according to claim 39, further comprising:

a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

41. The apparatus according to claim 39, further comprising:

a maximum link depth designating unit which designates a maximum link depth; and

a searching unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a

range of said designated maximum link depth.

42. The apparatus according to claim 41, further comprising:  
a searching unit which discriminates an internal link and an  
5 external link on the basis of the document address of the related document and  
excludes the documents of the external link from the targets of the information  
extraction.

43. The apparatus according to claim 41, further comprising:  
10 a searching unit which executes the information extracting  
process in order of the document in which a value of the link depth is small.

44. The apparatus according to claim 43, further comprising:  
a searching unit which discriminates an internal link and an  
15 external link on the basis of the document address of the related document and  
excludes the documents of the external link from the targets of the information  
extraction.

45. The apparatus according to claim 39, wherein said related  
20 document includes at least one of a link destination document, a link source  
document, and an upper document of the target document.

46. The apparatus according to claim 45, wherein said upper  
document is at least either a document of a specific name existing in a one-  
25 upper directory of the target document or a link source document existing in  
the one-upper directory.

47. The apparatus according to claim 39, further comprising:  
a category designating unit which designates a category of the  
information to be extracted; and  
a searching unit which extracts the information belonging to the  
category designated by said category designating unit.

48. The apparatus according to claim 47, further comprising:  
a searching unit which discriminates an internal link and an  
external link on the basis of the document address of the related document and  
excludes the documents of the external link from the targets of the information  
extraction.

49. The apparatus according to claim 47, further comprising:  
a maximum link depth designating unit which designates a  
maximum link depth; and  
a searching unit which, in the case where the information could  
not be extracted from the target document, recursively executes a process for  
extracting the information from the related document of said document in a  
range of said designated maximum link depth.

50. The apparatus according to claim 49, further comprising:  
a searching unit which discriminates an internal link and an  
external link on the basis of the document address of the related document and  
excludes the documents of the external link from the targets of the information  
extraction.

51. The apparatus according to claim 49, further comprising:

a searching unit which executes the information extracting process in order of the document in which a value of the link depth is small.

52. The apparatus according to claim 51, further comprising:

5 a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

10 53. The apparatus according to claim 47, wherein said related document includes at least one of a link destination document, a link source document, and an upper document of the target document.

15 54. The apparatus according to claim 53, wherein said upper document is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory.

20 55. The apparatus according to claim 47, further comprising:

a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure; and

25 a searching unit which, in the case where an extraction result of an upper layer is missing only in an extraction result of a lower layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result of said lower layer from the related document of

said target document, and outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer.

5        56.        The apparatus according to claim 55, further comprising:  
              a searching unit which discriminates an internal link and an  
external link on the basis of the document address of the related document and  
excludes the documents of the external link from the targets of the information  
extraction.

10        57.        The apparatus according to claim 55, further comprising:  
              a maximum link depth designating unit which designates a  
maximum link depth; and  
              a searching unit which, in the case where the information could  
15 not be extracted from the target document, recursively executes a process for  
extracting the information from the related document of said document in a  
range of said designated maximum link depth.

20        58.        The apparatus according to claim 57, further comprising:  
              a searching unit which discriminates an internal link and an  
external link on the basis of the document address of the related document and  
excludes the documents of the external link from the targets of the information  
extraction.

25        59.        The apparatus according to claim 57, further comprising:  
              a searching unit which executes the information extracting  
process in order of the document in which a value of the link depth is small.

60. The apparatus according to claim 59, further comprising:  
a searching unit which discriminates an internal link and an  
external link on the basis of the document address of the related document and  
excludes the documents of the external link from the targets of the information  
extraction.

61. The apparatus according to claim 55, wherein said related  
document includes at least one of a link destination document, a link source  
document, and an upper document of the target document.

62. The apparatus according to claim 61, wherein said upper  
document is at least either a document of a specific name existing in a one-  
upper directory of the target document or a link source document existing in  
the one-upper directory.

63. The apparatus according to claim 55, further comprising:  
a searching unit which has a predetermined synthesizing rule in  
the case of synthesizing a plurality of character strings expressed by the layer  
structure and forms a character string of a processing result in accordance  
with said synthesizing rule.

64. The apparatus according to claim 55, further comprising:  
a searching unit which forms a character string of a processing  
result by coupling a plurality of character strings in order from the extraction  
result of the upper layer to the extraction result of the lower layer on the basis  
of the layer structure.

65. The apparatus according to claim 64, further comprising:

a searching unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule.